

Six new recombinant inbred populations for the study of quantitative traits in *Arabidopsis thaliana*

Carmel M. O'Neill · Colin Morgan · Jane Kirby · Hendrik Tschoep · Polo Xiaoyi Deng · Mahon Brennan · Ulises Rosas · Fiona Fraser · Caroline Hall · Samantha Gill · Ian Bancroft

Received: 17 July 2007 / Accepted: 4 December 2007 / Published online: 12 January 2008
© Springer-Verlag 2008

Abstract Quantitative approaches are now widely used to study the genetic architecture of complex traits. However, most studies have been conducted in single mapping populations, which sample only a fraction of the natural allelic variation available within a gene pool and can identify only a subset of the loci controlling the traits. To enable the progress towards an understanding of the global genetic architecture of a broad range of complex traits, we have developed and characterised six new *Arabidopsis thaliana* recombinant inbred populations. To evaluate the utility of these populations for integrating analyses from multiple populations, we identified quantitative trait loci (QTL) controlling flowering time in vernalized plants growing in 16 h days. We used the physical positions of markers to align the linkage maps of our populations with those of six existing populations. We identified seven QTL in genomic locations coinciding with those identified in previous studies and in addition a further eight QTL were identified.

Communicated by A. Charcosset.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-007-0696-9) contains supplementary material, which is available to authorized users.

C. M. O'Neill · C. Morgan · J. Kirby · P. X. Deng · U. Rosas · F. Fraser · C. Hall · S. Gill · I. Bancroft (✉)
John Innes Centre, Norwich Research Park, Colney,
Norwich NR4 7UH, England
e-mail: ian.bancroft@bbsrc.ac.uk

H. Tschoep
Max-Planck-Institute for Molecular Plant Physiology,
14476 Potsdam-Golm, Germany

M. Brennan
Monsanto International, Rue des Vignerons 1A,
1110 Morges, Switzerland

Introduction

Natural genetic variation can be exploited to advance studies in many aspects of plant biology and is central to evolutionary biology and plant breeding (Borevitz and Nordborg 2003). Molecular dissection of the natural genetic variation controlling quantitative developmental traits is made difficult by their complex nature. Genes that contribute to variation in quantitative traits are difficult to identify as each locus is typically responsible for only a small proportion of the total phenotypic variation (Glazier et al. 2002). Such loci are termed quantitative trait loci (QTL) and they can be identified and mapped by quantitative genetic approaches. These provide estimates of the numbers of loci, the type and magnitude of their effects, interactions between genes and gene by environment interactions; collectively referred to as the genetic architecture of a trait. However, QTL can only be identified where the alleles present at a locus differ between the parents of the population. Consequently, the sets of QTL identified will often be population-specific (Lynch and Walsh 1998). Therefore, any description of the genetic architecture of a trait made from a single population, either natural or experimental, is likely to describe only a small part of the “global genetic architecture” for a trait within a species (Symonds et al. 2005). In order to understand the complete range of genetic variation that is responsible for the quantitative control of trait variation, many experimental segregating populations, derived from a diverse array of accessions, are needed (Tonsor et al. 2005).

In addition to the availability of segregating populations, QTL analyses require the construction of linkage maps for those populations. Sequence-based markers are particularly valuable in species for which complete genome sequences are available, such as *A. thaliana* (Arabidopsis Genome Initiative 2000). This is because they can be anchored to the

genome sequence, allowing precise alignment of linkage maps that have been constructed using differing sets of markers. Simple sequence repeat (SSR) and insertion/deletion (InDel) markers are two of the most widely used types of sequence-based markers. SSRs are sites where one or a few bases are tandemly repeated for a varying number of times. They are found in both coding and non-coding regions, highly polymorphic, co-dominant and abundant across the *A. thaliana* genome, and can be easily assayed by PCR (Casacuberta et al. 2000). There are many publicly available SSR markers for use in *A. thaliana* (<http://www.arabidopsis.org>; <http://www.arabidopsis.org/Cereon/index.jsp>; <http://www.inra.fr/vast/>). InDel markers are highly abundant, exist mostly in only two co-dominant variants and they are predominantly phenotypically neutral (Berger et al. 2001). Both SSR and InDel markers are suited to high throughput genotyping using capillary sequencers. Several genomic screens have been conducted for InDel and other sequence variations between *A. thaliana* accessions (Jander et al. 2002; Schmid et al. 2003; Nordborg et al. 2005), resulting in many publicly available markers (<http://www.arabidopsis.org/Cereon/index.html>; <http://www.mpiz-koeln.mpg.de/masc/>; <http://www.walnut.usc.edu/2010/>).

Arabidopsis thaliana is a largely autogamous species. Consequently, the most favoured approach to the development of “immortal” mapping populations has been the development of recombinant inbred (RI) populations. These are produced by single seed descent to at least the F8 generation, at which each line is practically homozygous and genetically identical individuals can be propagated indefinitely. This gives RI populations several important advantages over, for example, F2 populations. The principle advantages are that extensive linkage maps can be constructed and the lines can be analysed, in replicated experiments, for a wide range of phenotypic characteristics. The research potential of RI populations for the study of complex traits in *A. thaliana* has been demonstrated by the two most frequently used populations Ler × Col (Lister and Dean 1993) and Ler × Cvi (Alonso-Blanco et al. 1998). Both have been publicly available for a number of years and in that time the Ler × Cvi population has been analysed for in excess of forty traits (Koornneef et al. 2004). A recent example of how this type of population can be exploited was demonstrated by Pfalz et al. (2007) when they analysed a new RI population, Da(1)-12 × Ei-2, for resistance to *Pieris brassicae*. Currently, nine RI populations are available from the Arabidopsis Biological Resource Centres (<http://www.arabidopsis.info/>). Further RI populations are presently being developed (<http://www.inra.fr/vast/>). The parental accessions used to generate many of the RI populations have been included in several extensive genotyping studies (Nordborg et al. 2005; El-

Lithy et al. 2006; Schmid et al. 2006; Clark et al. 2007), which aimed to explore the phylogenetic relationships of a broad range of accessions.

Arabidopsis has a wide distribution throughout the northern hemisphere and grows in a range of environments. It is therefore an excellent system to use in understanding the molecular variation underpinning adaptive variation. The timing of the transition from vegetative to reproductive development has significant consequences for the reproductive success of plants and is the result of the integration of both environmental and endogenous cues. A complex regulatory network has evolved, consisting of multiple pathways that quantitatively regulate a set of genes (the floral pathway integrators) whose activity causes the transition of the meristem to reproductive development (Simpson and Dean 2002; Henderson and Dean 2004; Sung and Amasino 2005). Two particularly important components of this network are the genes at the FRI (FRIGIDA) and FLC (FLOWERING LOCUS C) loci. The range of natural variation in these genes was investigated using many different *Arabidopsis* accessions (Gazzani et al. 2003; Shindo et al. 2005, 2006). Another important gene involved in the control of flowering time *MAFI* (*MADS AFFECTING FLOWERING 1*) has also been found to be a target of natural selection (Werner et al. 2005). Light has a major role in modulating flowering time and allelic variation at the locus for the photoreceptor PHYC (PHYTOCROME C) was shown to affect this trait (Balasubramanian et al. 2006). Investigations of the genetic control of flowering time have been conducted using a range of RI populations (Lister and Dean 1993; Alonso-Blanco et al. 1998; Loudet et al. 2002; El-Lithy et al. 2004, 2006; Werner et al. 2005) and differing numbers of QTL identified. Many, but not all, of these co-localise in different populations.

We report the development and characterisation of six new RI populations, each comprising 94 independent lines. We also demonstrate their applicability to QTL mapping, using flowering time as an example. The populations were developed on the basis of the parent accessions differing in seed lipid composition and representing a broad range of geographic origins (O'Neill et al. 2003). The populations have been deposited with the Arabidopsis Biological Resources Centres (<http://www.arabidopsis.info/>) for public distribution (catalogue numbers to follow).

Materials and methods

Generation of the recombinant inbred populations

Seeds for each of the accessions used for the development of the populations were received from the Nottingham Arabidopsis Stock Centre (NASC; <http://www.arabidopsis.info/>).

Their geographic origins and catalogue numbers, the combinations crossed and population names are listed in Table 1. The line JIC240#14 was isolated as an apparent contaminant in the seeds we received of the Mz-0 accession. Fingerprinting all 364 available accessions for a range of markers revealed no matches, therefore the line has been submitted to NASC as a new accession (catalogue number to follow).

Before the crosses used to generate the RI populations were made, a single plant from each accession was grown and selfed. These selfed seed stocks were used for crossing, by standard emasculation procedures, to produce unidirectional F1 plants. The F1 plants were confirmed to be hybrids by molecular marker analysis and selfed. The RI populations were developed from the F2 lines by single seed descent to F7. The process was accelerated and the likelihood of crossing reduced, by removing side branches. Plants of the F8 generation were allowed to fully develop. Plastic sleeves were placed around the individual plants to prevent cross fertilisation and facilitate seed collection.

Plant growth conditions for RI population generations

Seeds were sown on to Arabidopsis soil mix: Levington's M2 compost with 4 mm grit (8:1) and intercept (to protect against aphids) in P40 trays (O'Neill et al. 2003). Plants were vernalized at the seed stage at every generation. To achieve this, they were placed, directly after sowing, in a 5°C chamber with an 8 h photoperiod for 6 weeks. After this time, the trays were moved to a temperature-controlled glasshouse at 22°C (day) and 18°C (night) temperature, with supplementary sodium lighting to provide a 16 h photoperiod.

Plant growth conditions for flowering time analysis

F8 seeds were sown to Arabidopsis soil mix and vernalized for 6 weeks as described above. The plants were then moved to a controlled environment room (CER) or a glasshouse (GH). Both provided a 22°C day/18°C night, 16 h photoperiod with supplementary light provided by sodium lamps. When the first true leaves were fully developed,

seedlings were pricked out to Arabidopsis soil mix in 7 cm pots. Flowering time was recorded as the number of rosette leaves present when the bolt had extended 1.5 cm. All populations were grown in replicated experimental designs. Three of the populations SG, KB and TJ were grown under GH conditions in a complete randomised block design. The plant positions within trays were randomised using GenStat (Payne 2000). Three replicate plants were observed per line. Two of the populations CA and WC were grown in the CER as complete randomised blocks with two replicate plants. The blocks were split into two equal sub-blocks and grown in two separate occasions. The sixth population, NG, was grown under CER conditions as a complete fully randomised block with two replicate plants. To minimise positional effects, plant trays were rotated 180° and moved along the bench, one tray position each week.

Marker development and assay

The markers used in this work were identified from four different resources: 48 SSRs were identified from the TAIR database (<http://www.arabidopsis.org>), 28 InDels were selected from the Monsanto database (<http://www.arabidopsis.org>). Another 37 markers were developed at the John Innes Centre for this work. Twenty-three of these were SSRs identified by using repeat sequences in BLASTN (<http://www.blast.wustl.edu>) alignments against the Arabidopsis genome sequence. The remaining 14 were InDel markers, which were identified after re-screening the raw sequence data generated in the Nordborg SNP generation/sequencing programme (<http://www.walnut.usc.edu/2010/>) for InDels. The primer combinations used for the 37 new markers developed are shown in Table 2. The complete marker set used for the individual populations is available as supplementary materials online, or from <http://www.jic.bbsrc.ac.uk/staff/ian-bancroft/> (Supplementary Table 1).

Genomic plant DNA was extracted using the high throughput Qiagen Dneasy 96 preparation method by following the manufacturer's instructions (Qiagen GmbH, Hilden, Germany). The markers were amplified using PCR and the products were resolved on an ABI 3700 capillary sequencer in order to determine allele sizes (Genome Labo-

Table 1 The Arabidopsis accessions used to generate the six RI populations and their origins

Population	Female parent	Cat. No.	Origin	Male parent	Cat. No.	Origin
WC	Wt-5	N1612	Germany	Ct-1	N1092	Italy
SG	Sorbo	N931	Tadjikistan	Gy-0	N1216	France
KB	Kondara	N916	Tadjikistan	Br-0	N994	Czech
CA	Cvi-0	N902	Cape Verde Islands	Ag-0	N936	France
TJ	Ts-5	N1558	Spain	240#14	N9623	Unknown
NG	Nok-3	N1404	Netherlands	Ga-0	N1180	Germany

Table 2 The 37 newly developed SSR and InDel markers

Marker	Chromosome	Position (Mb)	Forward primer 5'-3'	Reverse primer 5'-3'	Annealing temperature
jconn1_0.8	1	0.8	AAGATTGTATGGTTTTTGAGGATGAC	TCAGGATTTGCTTGTTTTTATATCTG	55
jconn1_2.7	1	2.7	ATGACGAGACAAAAGATAGAGAAAGC	AGTCATCGCCATTTTCAAAGATA	55
jconn1_2.7x	1	2.7	ACGAGACAAAAGATAGAGAAAGCAAA	AAACAGAGCCTTAGTAAAGAGATCA	55
jconn1_3	1	3	CGAAACCTATTACACACAAAATAGAGA	CAAAAGGTCACTACTTTGAGAAACAA	55
f8k728985	1	7.6	GGACAACAGCAAATGCCTCT	GAGGGAGTCCCTAGCAGAT	50
jconnchr1_9.9	1	9.9	TTTCATCGTACTTGTAGTTGTGTTG	CACATAACGGTTATAACAATGAGGAGTT	55
jconnchr1_11	1	11	TATGTAACCTATGTTGAAGAGGGGAAA	GCGATCAACAACCTATAGAACATTTTAC	55
t16o9 68289	1	12.7	TGGAAGAGAGAGGGCGCTAC	GTGCTTCATCCCTGCAAAAGT	50
f14o426482	1	13.2	TCGGAAAGCAATTTGGTACCTC	TTCTCGCCGATAAAGTCTGTTT	50
f19c2496807	1	18.7	CAACATGAAATATGTGACTTTGGA	TTTTTCGGGTGAAGAAATCAA	50
f5i1449495	1	24	CTGCCGTGAAAATGTCGAAAC	GGCATCACAGTTCTGATTC	50
f5a1859436	1	26.2	TCCACACCCCAATGCTCTAA	TCCCTCGGAGACATTTCAAG	50
jconn2_0.7	2	0.7	ATAAATCCCCCTTTTAGTCTTGT	CTGTTGAAAGGGAAGAGATATTACGA	55
t16b233819	2	1.4	TTTTTACCCCGAATAATCGAC	GGGTAATATGTGAAACACCATGA	50
jconn2_3	2	3	TATATTGGTTATTTGGTTTGTCTCTG	ATGTTTTATTGATGCTCCCATTTGTG	55
jconnchr2_3	2	3	TTATAATTGGTTATTTGGTTTGTCTCTG	ATGTTTTATTGATGCTCCCATTTGTG	55
t12j241450	2	3.6	TTCCGATCAAAAAGATGTGTG	TGCATCTTGAATTTGATTTGGA	50
jconn2_5.2	2	5.2	ACAAGGTCAAAAAGAGTAACTAAGGG	TTTCTCAGACATTAGACCTTAGATCG	55
jconnchr2_5.5	2	5.5	GAATCTGCTTAATTTGTTTATGATGTC	ATGTGATGATGAACTAAGATTTGATG	55
t32f646516	2	13.8	TGTGGTCCAAACACTCTC	TTGGAATTTGGAGGATCACCT	50
f17a93 1902	3	2.1	AGAGAGACAGGGCAGCAAAAC	ATTGCTTTTGTGGCGCTTAC	50
msd2129380	3	7.7	TTCCGATTTTATGATTAATAACTATGG	GAAATATAAGATGGCAAAATGAAGTT	50
t13j1016610	3	11.5	AGGGCTTACTAGGGATTG	CGTTAAACGGGCGTTACAGAA	50
t18d1241525	3	15.4	AGATGAGCCTTGTGGATGG	CATGTTGGCTTTTGTGAAGAAAG	50
t16k521877	3	18.9	TGTCGAAAATAAAAATGACCGTTA	AGCTGATGAGTGGATGTGGA	50
f14g 1617354	4	4.1	GGGTTCTGGAGTTTGGTGTG	TAGATGTTGGCGTTGTGCTT	50
f27g 1959898	4	12.8	CATCATCCTTTTGCCCTCTC	GGCCTGCAATTTGAGACCTA	50
t5j1748070	4	17.7	TGACGGCAGATTCAGAGAGA	AGGGAGGACGAAGAATGAGG	50
t20l159379	5	0.3	AAATATCCGAACCTTGCATTG	CAAAAATCCGCACACATAAT	50
f14f1847646	5	3.8	TCGAGCGGAATACAGGTACA	TCATTAATGCATTTGCGAAAAA	50
jconnchr5_7.8	5	7.8	ATTAGCTTCTCTCTCTCTCTCTC	ACTGATGGTTGTGTTGCGATTAAG	55
t15f177050	5	12.2	ACCTCGATTTTCACTACTTTTAGC	TTTTTGCTCCGAATGCTTCA	50
jconnchr5_12.7	5	12.7	AGAAACTTCACATGGTAAGAACTCTC	CAAAATAAACGACTACGAGAACAAAGAG	55
jconnchr5_13.6	5	13.6	AAGTATCCAGAAAAGAGAAAGAAAG	CTTTGGCTAAAGAGTAAATAACCAAGTTTC	55
mql58836	5	19.2	TTTGTGTGATAAATGGAGTTTCTGA	TTTACCTACCAAAAATTCATCAATCA	50
mrg2138439	5	25.2	ACCTAGACCACCAACCACCA	CAAAATTTGCGAAATTTGATCG	50
k8a1022396	5	26.8	CCTACTGGCAGCCAAATATC	GGAAAAAGATGATCGGAGGA	50

ratory Services, John Innes Centre). Markers identified as being polymorphic between the parents of a particular population were used to genotype the 94 individuals in each population.

Genetic mapping and QTL analysis

Genetic linkage maps for each recombinant inbred population were constructed using the Kosambi mapping function in the JoinMap program version 3 (Van Ooijen and Voorrips 2001) based on the polymorphic marker data generated for each population. The segregation pattern of parental alleles was determined for all loci on each linkage group using the JoinMap program. The significance values of the segregation ratios for all marker alleles were tested using a χ^2 test.

Flowering time was analysed separately for each experiment by analysis of variance using GenStat and the mean line values used for QTL mapping of each population independently and to estimate heritabilities (h^2_b). To identify the regions of the genome containing QTL, we initially used interval mapping as implemented by the software package Map QTL version 4 (Van Ooijen et al. 2002). These were further analysed using composite interval mapping (MQM) using markers identified as cofactors by the automatic cofactor selection function in MapQTL. Genome wide significance thresholds of 5% for the LOD values were determined using the permutation command within MapQTL. The number of permutations was set at 10,000.

Results

Generation of the RI populations

Six new RI populations derived from the twelve accessions described in Table 1 were developed by single seed descent to the F8 generation. To reduce the effect of the segregation for late flowering and seed dormancy, all lines, at each generation, were vernalized at 5°C at the seed stage for 6 weeks. Each population consisted of 94 lines, which are

available from the Nottingham Arabidopsis Stock Centre (NASC; <http://www.arabidopsis.info/>). We refer to the populations by names derived from the first letters of the name of the parental accessions, as listed in Table 1.

Linkage map development

Linkage maps are required for populations if they are to be used for QTL analysis. We developed a panel of 113 polymorphic molecular markers, derived from four different resources, as described in the material and methods section. Details of this marker set, and the linkage maps constructed for the six populations, are available as supplementary materials online and from <http://www.jic.bbsrc.ac.uk/staff/ian-bancroft/>. The linkage maps for the individual populations consisted of between 47 and 58 markers (Table 3). The maps ranged in length from 297 cM for SG to 470 cM for the WC population (Table 3). These are comparable to the lengths of linkage maps reported for other RI populations (Alonso-Blanco et al. 1998; Loudet et al. 2002; El-Lithy et al. 2006). All of the markers used in the genetic mapping were collinear with their expected physical positions, i.e. relative to the genome sequence. The linkage maps developed had average distances of 6.2–9.6 cM between markers (Table 3). However, all populations contained larger intervals. The maximum interval in CA, 41 cM, was found on chromosome 1 between the markers at 66 and 107 cM. In the TJ population, the largest interval, 38 cM, was found between 19 and 57 cM on chromosome 3. That in the WC population was located on chromosome 1 between 49 and 81 cM and a similar sized gap was also found on chromosome 1 in the KB population between 33 and 61 cM. In NG, the largest gap occurred on chromosome 5, between 33 and 55 cM and in SG the largest gap was at the top of chromosome 5, between 0 and 18 cM.

The ratio of physical distance (kb) to genetic distance (cM) provides a measure of the frequency with which recombination events have occurred during the development of the populations. The data given in Table 3 represents the whole-genome mean ratios for each of the populations. The overall ratios for the populations varied

Table 3 Genetic analysis of the recombinant populations at the F8 generation

Population	Total Mb	Total cM	Average distance between markers (cM)	Number of markers used in the maps	kb/cM	Residual heterozygosity (%)	Maternal alleles in the populations (%)
NG	114	448	7.7	58	255	1.0	45.6
TJ	108	386	7.3	53	280	0.8	49.2
CA	107	453	9.6	47	236	0.65	40.5
KB	104	351	6.8	52	296	1.0	41.4
SG	97	297	6.2	48	326	0.8	43.9
WC	112	470	8.2	57	238	0.6	52.8

from a value of 326 Kb/cM for SG to a value of 236 Kb/cM for CA. These data could be indicative of variation in recombination rate during the development of the different RI populations. It has been reported that structural chromosome variants between accessions are a common occurrence (Koornneef et al. 2003) and cytogenetic studies have suggested the existence of differences in local recombination rates (Sanchez-Moran et al. 2002). However, to investigate this in the populations generated here, a much larger number of markers would be required. It should be noted that extensive suppression of recombination in parts of a genome will affect the ability to conduct map based cloning in such regions.

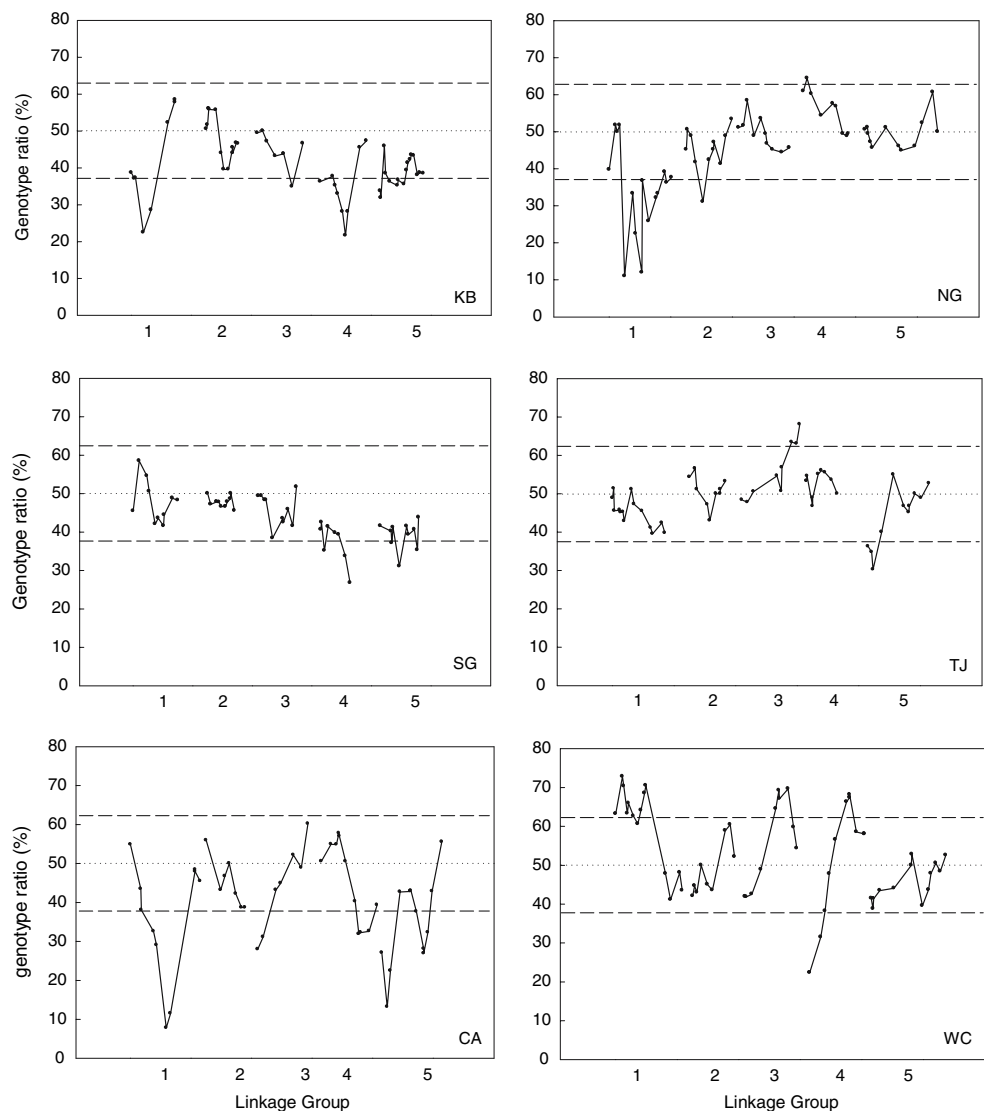
The proportion of the expected ca. 120 Mb genome represented by the linkage maps was calculated, based upon the positions of the terminal markers on each linkage group. Only the map for SG, which spanned 97 Mb, failed to represent at least 85% of the genome. This shortfall was

largely because we were unable to identify polymorphic markers at the distal ends of chromosomes 1 and 5.

The average frequencies of heterozygosity for all alleles in the six populations described were examined and found to be very close to 1.0% or less (Table 3). It is expected that by the F8 generation RIs will be essentially homozygous as the theoretical chance of any given marker being heterozygous is only 0.78% (Alonso-Blanco et al. 1998).

Where, no bias has occurred in the selection of individual plants during the development of a population, it is expected that the segregation ratio for each marker would be 1:1 for the parental alleles. Figure 1 presents the allele segregation ratios, with the female allele expressed as a percentage of the male and female alleles combined, for the six populations we developed. Significant distortion from the expected 1:1 at $P < 0.001$ was observed in all of the populations, as summarised in Supplementary Table 2.

Fig. 1 Allele segregation ratios for the six populations. The female alleles are expressed as a percentage of the male and female alleles combined. The significance threshold $P < 0.001$ (i.e. between 38 and 62%) is indicated by the horizontal dashed lines



There were regions of very unbalanced allele inheritance in populations CA, WC, KB and NG. In the KB population this was found on the linkage groups 1, 4 and 5, where in all cases the Br-0 alleles were in excess. The most significant deviation was in the interval 5–33 cM on chromosome 1, with the peak at the *aths0392* marker, where only 22% of the alleles were Kondara in origin. A similar ratio was found at the marker *t6k21ind15-15* on chromosome 4 in a region spanning 25–46 cM. The NG population displayed a very marked distortion on chromosome 1, where only 11% of the alleles for the *f9hind26-26* marker were from Nok-3. This was part of the region, 25–79 cM, in which all of the markers displayed distorted inheritance. The Nok-3 alleles were under represented at this location and also on chromosome 2 at *pls1*, however, the reverse was found on chromosome 4 at *t19j18ind30-30* (6 cM), where the Nok-3 was in excess. In CA, unbalanced allele inheritance was found on four of the linkage groups with the most severe being on 1 and 5 at the markers *jconnchr1_9.9* and *f14f1847646* with only 8 and 13%, respectively, of the alleles coming from Cvi-0. In all cases, the Ag-0 alleles were in excess. The region affected on chromosome 1 lay between 38 and 66 cM and those on chromosome 5 lay between 0 and 30 cM and 69–76 cM. In WC, unbalanced segregation ratios were observed on three of the chromosomes 1, 3 and 4. The distorted regions on chromosome 1 lay between the markers *nga59-f8k728985* (0–29 cM) and *aths0392-f1404682* (41–49 cM). On chromosome 3, the region between the markers *mzn14ind29-29-t16k521877* (50–70 cM) displayed allele distortion. On both chromosomes 1 and 3, allele inheritance was biased in favour of Wt-5. Two separate regions occurred on chromosome 4, 0–18 and 60–65 cM. In the top region the *ju30-31* marker had predominantly Ct-1 alleles (78%) but the reverse was the case at the *f27g1959898* marker (68%) at 65 cM. Some of the regions with distorted allele inheritance at least partially overlapped between some of the populations. This was found to be the case for chromosome 1 between CA, WC, KB and NG. Another region on chromosome 4 was common to both CA and WC. A region identified as having unbalanced allele composition towards the top of chromosome 5 in CA was also picked up in KB and TJ. Other regions identified as having distorted allele inheritance appeared to be population specific (Fig. 2).

On examination of the total number of alleles mapped in each population it was found that an unbalanced pattern of inheritance, as tested by a χ^2 test ($P < 0.01$), occurred in two of the populations, KB and CA. The maternal parents in these crosses were Kondara and Cvi-0, respectively, and accounted for only ca. 41% of the total alleles studied in those populations. The levels of the maternal progenitor genome in the other four crosses were closer to the expected value of 50% (Table 3).

Mapping QTL controlling flowering time

Under the conditions used in our experiments, subsets of lines in all the populations flowered at times beyond the ranges of their parental accessions, a phenomenon termed transgressive segregation. For example, the accessions Kondara and Br-0 bolted after 11 and 10 rosette leaves, respectively, while the individual RI lines in the KB population displayed leaf numbers ranging from 7 to 17. Flowering time is known to have a large genetic component, making it a highly heritable trait. This was found to be the case in five of our six populations, where the values for heritability (h^2_b) of flowering time varied between 0.6 and 0.8 (Table 4). However, in the case of CA, it was very low at 0.2. This was most likely the result of the variation between replicates being relatively high.

Our analyses revealed between 1 and 5 QTL in each population, as summarised in Table 4. Significance for the LOD values, estimated through permutation testing, indicated that nine of these were significant at $P < 0.001$ (LOD values > 3.8), four were significant at $P < 0.01$ (LOD values > 3) and two were significant at $P < 0.05$ (LOD values > 2.3). Individually, these 15 QTL accounted for between 9 and 58% of the phenotypic variation for flowering time found in the populations. Only one QTL, SG.1, (LOD = 3.8) was detected on chromosome 1 this was positioned around the marker *t12p18ind8-8* at 74.5 cM and accounted for 14% of the phenotypic variation. The paternal Gy-0 allele was responsible for increasing the flowering time at this QTL. Similarly, just one QTL was found on chromosome 2. This QTL, KB.1, had a LOD value of 13 ($P < 0.001$) at 30.7 cM and accounted for 58% of the phenotypic variation in that population. In this case the allele responsible for delaying flowering time arose from the maternal parent Kondara. No QTL was found on chromosome 3 in any of the populations. In four of the populations NG, TJ, CA and SG, QTL were detected at the top of the upper arm of chromosome 4. Four QTL in the different populations mapped between 0.5 and 14 cM, and each accounted for between 9 and 38% of the phenotypic variation.

Of all of the chromosomes, the largest number of QTL, eight, were found to map to chromosome 5 and were distributed along its length. These QTL varied in size but the three strongest examples were found to map to the lower arm in both the WC (WC.1, WC.2) and TJ (TJ.3) populations. WC.1 and WC.2 had LOD values of 8.9 and 10.2, respectively, and each accounted for 37 and 41% of the variation found. The 2 LOD intervals for WC.1 lies between 93 and 103 cM, whilst WC.2 was lower down at 107 cM through to the end of that chromosome. The increasing alleles at both loci came from the maternal parent Wt-5. The QTL TJ.3 had a LOD score of 7.4 and the 2 LOD intervals for this QTL spanned from 74 to 97 cM. TJ.3 accounted for

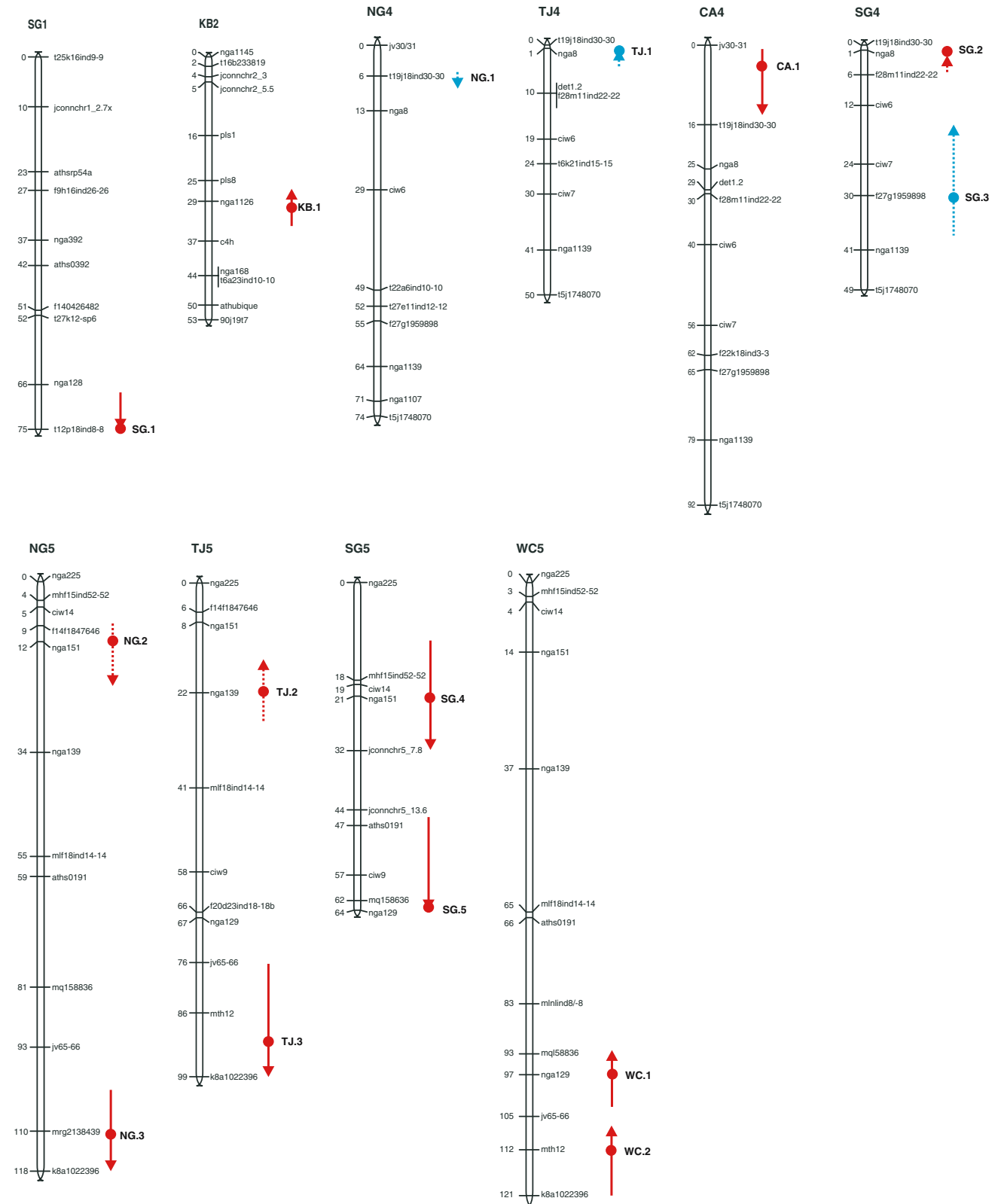


Fig. 2 QTL for flowering time for 6 RIL Arabidopsis populations. The position of the QTL are represented as 2 LOD intervals where sufficiently large, *solid lines*, otherwise the 5% significance LOD threshold is used, *dashed lines*. The position of peak LOD scores are indicated by *closed circles*. QTL detected by interval mapping are

indicated in *red*, while QTL detected only by CIM mapping are indicated in *blue*. *Upward arrows* indicate the maternal parent as the source of the increasing allele, *downward arrows* indicate the paternal parent as the source of the increasing allele

Table 4 QTL affecting flowering time in the six populations and heritability in the individual populations

Population	Heritability (h^2_b)	QTL analysis							
		QTL	Chr	LOD max	QTL position (cM)	Peak max (cM)	Variation explained (%)	Increasing allele	
KB ^a	0.69	KB.1	C2	13	26.5–34	30.7	58	Kondara	
NG ^b	0.65	NG.1 ^c	C4	2.5	6	6	10	Ga-0	
		NG.2	C5	4	7.7–20	11	19	Ga-0	
		NG.3	C5	4.7	101–117	110	22	Ga-0	
TJ ^a	0.63	TJ.1 ^c	C4	2.9	0–2	0	9	Ts-5	
		TJ.2	C5	3.1	13.7–26	19.7	18	Ts-5	
		TJ.3	C5	7.4	74–97	89.9	37	240#14	
CA ^b	0.2	CA.1	C4	7.4	0.5–14	4	38	Ag-0	
SG ^a	0.82	SG.1	C1	3.8	68–74.5	74.5	14	Gy-0	
		SG.2	C4	3.4	0–6	1.4	16	Sorbo	
		SG.3 ^c	C4	3.8	16–38	30.5	15	Sorbo	
		SG.4	C5	4.3	10–32	21.3	20	Gy-0	
		SG.5	C5	5.1	45.6–63.8	63.8	23	Gy-0	
WC ^b	0.6	WC.1	C5	8.9	93–103	96.8	37	Wt-5	
		WC.2	C5	10.2	107–end	112.2	41	Wt-5	

^a Population grown in a glass-house

^b Population grown in a controlled environment room

^c QTL detected only in the MQM analysis

37% of the phenotypic variation in this population. The 240#14 allele was the increasing allele in this cross. A second but weaker QTL also mapped in the TJ population to chromosome 5 but to the upper arm at 13.7–26 cM. This QTL, TJ.2, was a weaker QTL with a LOD score of 3.1

Analysis with CIM showed that when marker nga8 (SG.2) was used as a cofactor, SG.1 increased in significance from LOD 2.4 to 3.8 and when marker t12p18ind8-8 (SG.1) was used as a cofactor, SG.2 increased from LOD 3.4 to 4.5. Similar effects were observed for QTL NG.1 and NG.3, and TJ.1 and TJ.3. In the case of NG.1, it just reached the 5% significance level, when mrg2138439 (NG.3) was selected as a cofactor, while NG.3 also showed increased significance, increasing from LOD 4.7 to 5.4 with t19j18ind30-30 (NG.1) as a cofactor. Using nga8 (TJ.1) as a cofactor, TJ.3 increased from LOD 7.4 to 8.7, while selecting both jv65-66 and k8a1022396 as cofactors (TJ.3), TJ.1 increased from non-significance to a LOD of 2.9. In SG, when nga151 (SG.4) was used as a cofactor, a further QTL not previously identified through interval mapping, and designated SG.3, was found with a LOD score of 3.8 between 16 and 38 cM. This QTL was not visible when any of the other QTL cofactors were examined.

Discussion

In this paper, we describe six new RI populations, which are communally available through the Arabidopsis Biological Research Centres. These populations will provide an additional resource for the large research community focussed on *A. thaliana* and support investigations on a

broad range of plant biology disciplines. Genetic polymorphisms between the parental genotypes are crucial for linkage mapping while allelic variation is necessary for QTL detection. In an effort to maximise genetic variation in the populations generated, the accessions combined in each of the six original crosses were from geographically diverse locations. Eleven of the accessions were screened in two large-scale studies (Nordborg et al. 2005; El-Lithy et al. 2006), which investigated the levels of polymorphism between a wide range of accessions and subsequently calculated their relatedness. From these data, it is clear that the accessions used in each of the parental pairing combinations, with the exception of TJ, could be discriminated. It therefore appears that the strategy behind the selection of suitable crosses was appropriate; geographical location can be used as a means to select for genetic divergence among accessions.

To determine the complete genetic architecture for any trait, it must be analysed in many different genetic backgrounds over many occasions under different environmental conditions (Symonds et al. 2005). If this objective is to be met access to a large number of “immortal” mapping populations such as RIs is necessary. The genetic linkage maps constructed for the six populations described here are comparable in length to previous ones described in the literature (Alonso-Blanco et al. 1998; Loudet et al. 2002; El-Lithy et al. 2006). Our aim was to develop maps with an optimal distance of ca. 10 cM between the markers. Though the average distance between markers in the populations ranged between 6.2 and 9.6 cM, there were also numerous gaps in each that exceeded this 10 cM value. Difficulties with the identification of polymorphic markers at the time

of the screens meant that some large gaps remain in the genetic maps generated. The presence of large intervals will hinder QTL mapping in these regions. However, with the advances in marker development, it should be possible for future users to identify suitable markers to fill in any gaps in a population map of interest. Despite this limitation, we have developed linkage maps that, although rudimentary, are fully anchored to the *A. thaliana* genome sequence. By keeping each population size to 94 lines, it enables high throughput molecular screening to be conducted. The markers used in this work routinely gave reproducible results.

In any QTL mapping work, the quality of the populations used is fundamental. The average frequencies of heterozygosity for all alleles in the six populations described here were found to be ca. 1.0% or less. This value is very much in line with the expectation that by the F8 generation RIs are essentially homozygous and there is only a 0.78% theoretical chance of any marker being heterozygous (Alonso-Blanco et al. 1998). The segregation ratios across the six populations were found to be largely below 1:2, though in each of the populations there were exceptions to this with some regions displaying severely distorted allele inheritance pattern. Such regions are likely to affect the detection of QTL, which lie within them. In four of the populations, CA, KB, WC and NG regions with biased segregation ratios for marker alleles on chromosome 1, at least partially overlapped between 2.8 and 23.5 Mb. A number of previous studies on different RI populations also found unbalanced inheritance in this region of the genome (Clerkx et al. 2004; El-Lithy et al. 2006; Pfalz et al. 2007). In work conducted by Alonso-Blanco et al. (2003) on the Ler × Cvi population a delay of germination (DOG) QTL was mapped onto chromosome 1 at 7.3 Mb. The Cvi allele at this locus increased the time of seed storage required to allow seed germination to occur. In the CA population, at marker position 9.9 Mb, we found that the level of Cvi-0 alleles dropped to 7.8% of the total alleles at that locus while Kondara contributed just 20% of the alleles at for the marker positioned at 10.9 Mb in the KB population. It was observed at the F3 generation that these two populations exhibited reduced germination rates compared to the other four populations with only 85 and 79%, respectively, of the seeds germinating. It would be expected that the 6-week vernalisation treatment would overcome any over-ripening requirements. However, these results suggest that this may not be the case.

Flowering time represents a suitable trait with which to test the suitability of mapping populations for QTL analysis, partly because many of the major genes involved in its regulation have already been identified. A number of Arabidopsis RI populations have previously been used to study flowering-time (Alonso-Blanco et al. 1998; Loudet et al. 2002; El-Lithy et al. 2004, 2006; Koornneef et al. 2004 for

review; Werner et al. 2005; Balasubramanian et al. 2006). Understanding the allelic variation at the loci identified and the subsequent effects on the flowering phenotype is now the major challenge (Gazzani et al. 2003; Shindo et al. 2006). By using markers physically anchored to the Arabidopsis genome we are able to compare the results for flowering time from our six new populations against the results generated in populations studied previously (Loudet et al. 2002; El-Lithy et al. 2004, 2006; Werner et al. 2005) as displayed in Fig. 3. The positions of seven QTL identified in this work coincided with QTL identified in earlier studies using different populations: Bay-0 × Shahdara (Loudet et al. 2002), Ler × Sha (El-Lithy et al. 2004), Ler × An-1, Ler × Kas-2, and Ler × Kondara (El-Lithy et al. 2006), and Nd-1 × Col (Werner et al. 2005). These previous studies were also carried out under long day conditions but unlike the work described in this paper the plant material was not exposed to a vernalization treatment. Of the seven QTL found to be in common, the first SG.1 was positioned around 24 Mb on chromosome 1 close to the position of the flowering locus T-gene at 24.3 Mb and *MAF1* at 28.8 Mb, which was detected in the Nd-1 × Col population (Werner et al. 2005). Another QTL found in common between the data sets was identified at the top of chromosome 4. This QTL was picked up in five of the previous populations and in the CA population in this work. It is possible to speculate that the gene responsible for the phenotype is *FRI*. Gazzani et al. (2003) showed that Cvi-0 carries an *FRI* allele with an in-frame translation stop codon within exon 1 and F1 analysis suggested that this allele is non-functional. We found that the allele responsible for increasing flowering time in the CA population came from Ag-0. On chromosome 5 both NG.2 and SG.4 were positioned around the nga151 marker at 4.7 Mb. This is very close to the flowering time genes *CO* at 5.1 Mb, *FY* at 4.3 Mb and *FLC* at 3.2 Mb. *FLC* has been shown to be involved in the vernalisation pathway controlling flowering time and given that all the material used in this work was vernalised, there is a reduced chance that this is the gene responsible for these QTL. QTL in this region were also found in the Ler × An-1, Ler × Kas and Ler × Kon populations. Large QTL were identified in the TJ and WC populations towards the end of chromosome 5 at ca. 20 and 24 Mb. QTL in this region were also detected in the Ler × Sha (El-Lithy et al. 2004) and the Ler × Kondara and Ler × Kas populations (El-Lithy et al. 2006). In addition highly significant QTL were also identified in the NG and SG populations in this region. Major genes controlling flowering time in this region of the genome are *LFY* and the *MAF* genes 2, 3, 4 and 5.

Of the QTL unique to the new populations, the most significant one was picked up on chromosome 2 in the KB population at 11.6 Mb. This QTL accounted for 58% of the phenotypic variation with a LOD value of 13. A possible

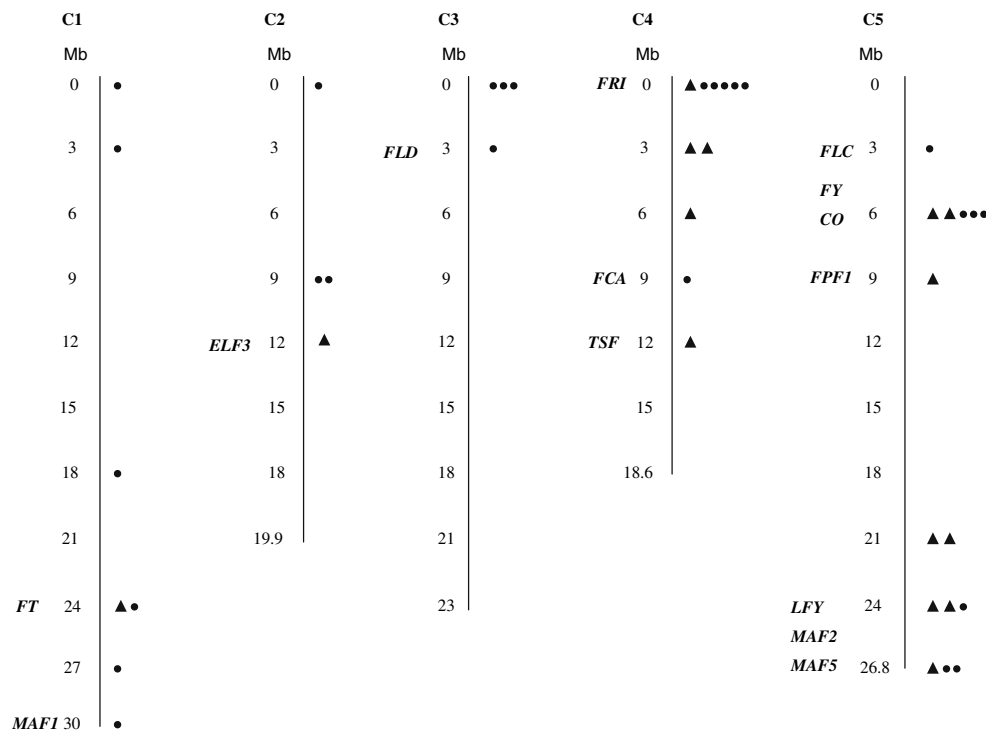


Fig. 3 Physical positions of QTL affecting flowering time across 12 populations. QTL identified in the populations reported here are represented by *filled triangle*. QTL identified in the populations studied by

Loudet et al. (2002), El-Lithy et al. (2004, 2006) and Werner et al. (2005) are represented by *filled circle*

candidate for this QTL is *ELF3*, which controls rhythmic circadian outputs under constant light conditions and is found at 11 Mb.

In the SG population, SG.1 (on chromosome 1) spans the region containing *FT* and SG.3 (on chromosome 4) spans the region containing *TSF*, which is a homologue of *FT*. SG.3 was only detected when the marker selected as a cofactor for SG.4 was applied in the MQM mapping. This suggests that SG.4, which maps in the region of *CO*, was masking the effect of SG.3. It should be noted, however, that the genetic map for chromosome 1 in SG did not provide complete coverage to the end of the physical chromosome, where another gene involved in the control of flowering, *MAF1*, is located.

The experiments described in this work were carried out under a defined set of growing conditions: long days following 6-weeks vernalisation. It might therefore be expected that allelic differences in genes involved in the vernalisation pathway would not be detected. This appears largely to be the case, with one exception, CA.1, for which *FRI* is a possible candidate. It should be noted that for all of the QTL, any clearer identification of the candidate genes requires further mapping beyond the preliminary results that we have presented. The involvement of loci on chromosome 5 in the control of flowering seems to be extensive, with 15 of the 41 QTL identified in the 12 populations mapping to this chromosome (Fig. 3). For four of our six

populations, two QTL were detected on chromosome 5. It is impossible to assign candidate genes to these all of these QTL, but it does suggest that this chromosome deserves special attention.

In this work, we have developed and characterised six new *Arabidopsis thaliana* recombinant inbred populations and shown the utility of these populations for the study of the complex trait flowering time. By using the physical marker positions we aligned the linkage maps of our populations with those of six existing populations and were therefore able to compare QTL for flowering time across 12 populations. The ability to conduct such comparative work between immortal populations will provide the research community with a powerful tool to aid in the understanding of the global genetic architecture of a broad range of complex traits.

Acknowledgments We would like to thank Judith Irwin and David Laurie for constructive criticism of drafts of this manuscript. This work was funded by the John Innes Centre Competitive Support Grant.

References

- Alonso-Blanco C, Peeters AJM, Koornneef M, Lister C, Dean C, van den Bosch N, Pot J, Kuiper MTR (1998) Development of an AFLP based linkage map of LER, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a Ler/Cvi recombinant inbred population. *Plant J* 14:259–271

- Alonso-Blanco C, Bentsink L, Hanhart CJ, Blankestijn-de Vries H, Koornneef M (2003) Analysis of natural allelic variation at seed dormancy loci of *Arabidopsis thaliana*. *Genetics* 164:711–729
- Arabidopsis Genome Initiative (2000) Analysis of the genome of the flowering plant *Arabidopsis*. *Nature* 408:796–815
- Balasubramanian S, Sureshkumar S, Agrawal M, Michael TP, Wessinger C, Maloof JN, Clark R, Warthmann N, Chory J, Weigel D (2006) The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. *Nat Genet* 38:711–715
- Berger J, Suzuki T, Senti KA, Stubbs J, Schaffner G, Dickson BJ (2001) Genetic mapping with SNP markers in *Drosophila*. *Nat Genet* 29:475–481
- Borevitz JO, Nordborg M (2003) The impact of genomics on the study of natural variation in *Arabidopsis*. *Plant Physiol* 132:718–725
- Casacuberta E, Puigdomenech P, Monfort A (2000) Distribution of microsatellites in relation to coding sequences within the *Arabidopsis thaliana* genome. *Plant Sci* 157:97–104
- Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, Huaming C, Frazer KA, Huson DH, Schölkopf B, Nordborg M, Rättsch G, Ecker JR, Weigel D (2007) Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* 317:338–342
- Clerkx EJM, El-Lithy ME, Vierling E, Ruys GJ, Blankestijn-De Vries H, Groot SPC, Vreugdenhil D, Koornneef M (2004) Analysis of natural allelic variation of *Arabidopsis* seed germination and seed longevity traits between the accessions Landsberg *erecta* and Shaldara, using a new recombinant inbred line population. *Plant Physiol* 135:432–443
- El-Lithy ME, Clerkx EJ, Ruys GJ, Koornneef M, Vreugdenhil D (2004) Quantitative trait locus analysis of growth-related traits in a new *Arabidopsis* recombinant inbred population. *Plant Physiol* 135:444–458
- El-Lithy ME, Bentsink L, Hanhart CJ, Ruys GJ, Rovito D, Broekhof JLM, van der Poel HJA, van Eijk MJT, Vreugdenhil D, Koornneef M (2006) New *Arabidopsis* recombinant inbred line populations genotyped using SNPWave and their use for mapping flowering-time quantitative trait loci. *Genetics* 172:1867–1876
- Gazzani S, Gendall AR, Lister C, Dean C (2003) Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Physiol* 132:1107–1114
- Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. *Science* 298:2345–2349
- Henderson IR, Dean C (2004) Control of *Arabidopsis* flowering: the chill before the bloom. *Development* 131:52–535
- Jander G, Norris S, Rounsley S, Bush D, Levin I, Last R (2002) *Arabidopsis* map-based cloning in the post-genome era. *Plant Physiol* 129:440–450
- Koornneef M, Fransz P, de Jong H (2003) Cytogenetic tools for *Arabidopsis thaliana*. *Chromosome Res* 11:183–194
- Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu Rev Plant Biol* 55:141–172
- Loudet O, Chaillou S, Camilleri C, Bouchez D, Daniel-Vedele F (2002) Bay-0 × Shaldara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in *Arabidopsis*. *Theor Appl Genet* 104:1173–1184
- Lister C, Dean C (1993) Recombinant inbred populations for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *Plant J* 4:745–750
- Lynch M, Walsh B (1998) *Genetics and analysis of quantitative traits*. Sinauer, Sunderland
- Nordborg M, Hu T, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calebrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg N, Shah C, Wall J, Wang J, Zhao K, Kalbfleisch T, Schulz V, Kreitman M, Bergelson J (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol* 3:e196
- O'Neill CM, Gill S, Hobbs D, Morgan C, Bancroft I (2003) Natural variation for seed oil composition in *Arabidopsis thaliana*. *Phytochemistry* 64:1077–1090
- Payne RW (2000) *GenStat: the guide to GenStat*, 4th edn. VSN, Oxford
- Pfalz M, Vogel H, Mitchell-Olds T, Kroymann J (2007) Mapping of QTL for resistance against the crucifer specialist herbivore *Peris brassicae* in the new *Arabidopsis* inbred line population, Da(1)-12 x Ei-2. *PLoS ONE* 2(6):e578
- Sanchez-Moran E, Armstrong SJ, Santos JL, Franklin FCH, Jones GH (2002) Variation in chiasma frequency among eight accessions of *Arabidopsis thaliana*. *Genetics* 162:1415–1422
- Schmid K, Rosleff-Sorensen T, Stracke R, Torjek O, Altmann T, Mitchell-Olds T, Weisshaar B (2003) Large scale identification and analysis of genome-wide single nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. *Genome Res* 13:1250–1257
- Schmid KJ, Torjek O, Meyer R, Schmutz H, Hoffmann MH, Altmann T (2006) Evidence for large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. *Theor Appl Genet* 112:1104–1114
- Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C (2005) Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of *Arabidopsis*. *Plant Physiol* 138:1163–1173
- Shindo C, Lister C, Crevillen P, Nordborg M, Dean C (2006) Variation in the epigenetic silencing of FLC contributes to natural variation in *Arabidopsis* vernalisation response. *Genes Dev* 20:3079–3083
- Simpson GC, Dean C (2002) *Arabidopsis*, the Rosetta stone of flowering time? *Science* 296:285–289
- Sung S, Amasino RM (2005) Remembering winter: toward a molecular understanding of vernalization. *Annu Rev Plant Biol* 56:491–508
- Symonds VV, Godoy AV, Alconada T, Botto JF, Juenger TE, Casal JJ, Lloyd AM (2005) Mapping quantitative trait loci in multiple populations of *Arabidopsis thaliana* identifies natural allelic variation for trichome density. *Genetics* 169:1649–1658
- Tonsor SJ, Alonso-Blanco C, Koornneef M (2005) Gene function beyond the single trait: natural variation, gene effects and evolutionary ecology in *Arabidopsis thaliana*. *Plant Cell Environ* 28:2–20
- Van Ooijen JW, Voorrips RE (2001) *JoinMap* version 3.0: software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands
- Van Ooijen JW, Boer MP, Jansen RC, Maliepaard C (2002) *MapQTL* 4.0: software for the calculation of QTL positions. Plant Research International, Wageningen, The Netherlands
- Werner JD, Borevitz JO, Warthmann N, Trainer GT, Ecker JR, Chory J, Weigel D (2005) Quantitative trait locus mapping and DNA array hybridization identify an *FLM* deletion as a cause for natural flowering-time variation. *PNAS* 102:2460–2465